
ONE-CLASS DETECTION OF CELL STATES IN TUMOR SUBTYPES.

Journal: Pac Symp Biocomput

Publication Year: 2016

Authors: Artem Sokolov, Evan O Paull, Joshua M Stuart

PubMed link: 26776204

Funding Grants: Center of Excellence for Stem Cell Genomics - Stanford

Public Summary:

Personalized medicine seeks to make treatment recommendations based on genetic properties of a patient. To make such recommendations, computer models can be trained to accurately recognize molecular states of patients' cells. The traditional approach is to train dichotomous predictors that answer yes/no questions: Does the patient have a certain type of breast cancer? Will the patient's tumor stop growing when exposed to a particular drug? Distinguishing among such a series of dichotomies can paint a detailed picture of the patient's diagnosis, prognosis and treatment response and develop a treatment plan based on this picture. To teach a dichotomous predictor to answer yes/no questions, one has to show it training examples from both categories. A big challenge in clinical applications is that the "no" category may not be available or well-defined. In a realistic scenario, the predictor doesn't see examples of "yes" and "no", but rather of "yes" and "don't know". In this paper, we propose to replace dichotomous predictors with one-class detectors. Detectors learn from examples in the "yes" category only. Upon learning to recognize the specific state of cells, they can then be applied to detect the presence of the same state in new patients, scoring the tumors of those patients according to the strength of the signal. We demonstrate that one-class detectors are more accurate than their dichotomous counterparts when presented with a collection of "yes" and "don't know" examples. We then go on to derive one-class detectors for the major breast and bladder cancer subtypes and reaffirm the connection between these two tissues. Finally, we train a one-class detector to recognize the embryonic stem cell signal and use it to show that the signal is present in the most aggressive subtype of breast cancer, which suggests "stem-like" properties of cells in these tumors.

Scientific Abstract:

The cellular composition of a tumor greatly influences the growth, spread, immune activity, drug response, and other aspects of the disease. Tumor cells are usually comprised of a heterogeneous mixture of subclones, each of which could contain their own distinct character. The presence of minor subclones poses a serious health risk for patients as any one of them could harbor a fitness advantage with respect to the current treatment regimen, fueling resistance. It is therefore vital to accurately assess the make-up of cell states within a tumor biopsy. Transcriptome-wide assays from RNA sequencing provide key data from which cell state signatures can be detected. However, the challenge is to find them within samples containing mixtures of cell types of unknown proportions. We propose a novel one-class method based on logistic regression and show that its performance is competitive to two established SVM-based methods for this detection task. We demonstrate that one-class models are able to identify specific cell types in heterogeneous cell populations better than their binary predictor counterparts. We derive one-class predictors for the major breast and bladder subtypes and reaffirm the connection between these two tissues. In addition, we use a one-class predictor to quantitatively associate an embryonic stem cell signature with an aggressive breast cancer subtype that reveals shared stemness pathways potentially important for treatment.